Exact solution of site and bond percolation

on small-world networks

Cristopher Moore[1,2] and M. E. J. Newman[2]

[1]*Departments of Computer Science and Physics, University of New Mexico,*

Albuquerque, New Mexico 87131

[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*

Abstract

We study percolation on small-world networks, which has been proposed as a simple model of the propagation of disease. The occupation probabilities of sites and bonds correspond to the susceptibility of individuals to the disease and the transmissibility of the disease respectively. We give an exact solution of the model for both site and bond percolation, including the position of the percolation transition at which epidemic behavior sets in, the values of the two critical exponents governing this transition, and the mean and variance of the distribution of cluster sizes (disease outbreaks) below the transition.

In the late 1960s, Milgram performed a number of experiments which led him to conclude that, despite there being several billion human beings in the world, any two of them could be connected by only a short chain of intermediate acquaintances of typical length about six [1]. This result, known as the "small-world effect", has been confirmed by subsequent studies and is now widely believed to be correct, although opinions differ about whether six is an accurate estimate of typical chain length [2].

The small-world effect can be easily understood in terms of random graphs [3] for which typical vertex–vertex distances increase only as the logarithm of the total number of vertices. However, random graphs are a poor representation of the structure of real social networks, which show a "clustering" effect in which there is an increased probability of two people being acquainted if they have another acquaintance in common. This clustering is absent in random graphs. Recently, Watts and Strogatz [4] have proposed a new model of social networks which possesses both short vertex–vertex distances and a high degree of clustering. In this model, sites are arranged on a one-dimensional lattice of size L, and each site is connected to its nearest neighbors up to some fixed range k. Then additional links— "shortcuts"—are added between randomly selected pairs of sites with probability ϕ per link on the underlying lattice, giving an average of $\phi k L$ shortcuts in total. The short-range connections produce the clustering effect while the long-range ones give average distances which increase logarithmically with system size, even for quite small values of ϕ.

This model, commonly referred to as the "small-world model," has attracted a great deal of attention from the physics community. A number of authors have looked at the distribution of path lengths in the model, including scaling forms [5–7] and exact and mean-field results [8,9], while others have looked at a variety of dynamical systems on small-world networks [4,10,11]. A review of recent developments can be found in Ref. [12].

One of the most important consequences of the small-world effect is in the propagation of disease. Clearly a disease can spread much faster through a network in which the typical person-to-person distance is $O(\log L)$ than it can through one in which the distance is $O(L)$. Epidemiology recognizes two basic parameters governing the effects of a disease:

the *susceptibility*—the probability that an individual exposed to a disease will contract it—and the *transmissibility*—the probability that contact between an infected individual and a healthy but susceptible one will result in the latter contracting the disease. Newman and Watts [6] studied a model of disease in a small-world which incorporates these variables. In this model a randomly chosen fraction p of the sites or bonds in the small-world model are "occupied" to represent the effects of these two parameters. A disease outbreak which starts with a single individual can then spread only within a connected cluster of occupied sites or bonds. Thus the problem of disease spread maps onto a site or bond percolation problem. At some threshold value p_c of the percolation probability, the system undergoes a percolation transition which corresponds to the onset of epidemic behavior for the disease in question. Newman and Watts gave an approximate solution for the position of this transition on a small-world network.

In this paper, we give an exact solution for both site and bond percolation on small-world networks using a generating function method. Our method gives not only the exact position of the percolation threshold, but also the values of the two critical exponents governing behavior close to the transition, the complete distribution of the sizes of disease outbreaks for any value of p below the transition, and closed-form expressions for the mean and variance of the distribution. A calculation of the value of p_c only, using a transfer-matrix method, has appeared previously as Ref. [13].

The basic idea behind our solution is to find the distribution of "local clusters"—clusters of occupied sites or bonds on the underlying lattice—and then calculate how the shortcuts join these local clusters together to form larger ones. We focus on the quantity $P(n)$, which is the probability that a randomly chosen site belongs to a connected cluster of n sites. This is also the probability that a disease outbreak starting with a randomly chosen individual will affect n people. It is *not* the same as the distribution of cluster sizes for the percolation problem, since the probability of an outbreak starting in a cluster of size n increases with cluster size in proportion to n, all other things being equal. The cluster size distribution is therefore proportional to $P(n)/n$, and can be calculated easily from the results given in this
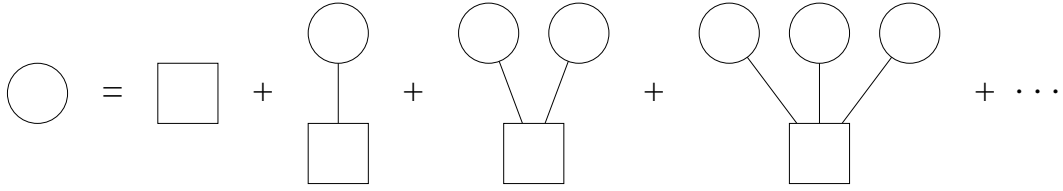
FIGURE 1 Graphical representation of a cluster of connected sites. The entire cluster (circle) is equal to a single local cluster (square), with any number $m \geq 0$ of complete clusters attached to it by a single shortcut.

paper, although we will not do so.

We start by examining the site percolation problem, which is the simpler case. Since $P(n)$ is difficult to evaluate directly, we turn to a generating function method for its calculation. We define

$$H(z) = \sum_{n=0}^{\infty} P(n)z^n. \tag{1}$$

For all $p < 1$, as we show below, the distribution of local clusters falls off with cluster size exponentially, so that every shortcut leads to a different local cluster for L large: the probability of two shortcuts connecting the same pair of local clusters falls off as L^{-1}. This means that any complete cluster of sites consists of a local cluster with $m \geq 0$ shortcuts leading from it to m other clusters. Thus $H(z)$ satisfies the Dyson-equation-like iterative condition illustrated graphically in Fig. 1, and we can write it self-consistently as

$$H(z) = \sum_{n=0}^{\infty} P_0(n)z^n \sum_{m=0}^{\infty} P(m|n)[H(z)]^m. \tag{2}$$

In this equation $P_0(n)$ is the probability of a randomly chosen site belonging to a local cluster of size n, which is

$$P_0(n) = \begin{cases} 1 - p & \text{for } n = 0 \\ npq^{n-1}(1-q)^2 & \text{for } n \geq 1, \end{cases} \tag{3}$$

with $q = 1 - (1-p)^k$. $P(m|n)$ is the probability of there being exactly m shortcuts emerging from a local cluster of size n. Since there are $2\phi k L$ ends of shortcuts in the network, $P(m|n)$ is given by the binomial

4

$$P(m|n) = \binom{2\phi kL}{m}\left[\frac{n}{L}\right]^m\left[1 - \frac{n}{L}\right]^{2\phi kL - m}.$$ (4)

Using this expression Eq. (2) becomes

$$H(z) = \sum_{n=0}^{\infty} P_0(n)z^n\left[1 + (H(z) - 1)\frac{n}{L}\right]^{2\phi kL} = \sum_{n=0}^{\infty} P_0(n)[ze^{2k\phi(H(z)-1)}]^n,$$ (5)

for L large. The remaining sum over n can now be performed conveniently by defining

$$H_0(z) = \sum_{n=0}^{\infty} P_0(n)z^n = 1 - p + pz\frac{(1-q)^2}{(1-qz)^2},$$ (6)

where the second equality holds in the limit of large L and we have made use of (3). $H_0(z)$ is the generating function for the local clusters. Now we notice that $H(z)$ in Eq. (5) is equal to $H_0(z)$ with $z \to ze^{2k\phi(H(z)-1)}$. Thus

$$H(z) = H_0(ze^{2k\phi(H(z)-1)}).$$ (7)

$H(z)$ can be calculated directly by iteration of this equation starting with $H(z) = 1$ to give the complete distribution of sizes of epidemics in the model. It takes n steps of the iteration to calculate $P(n)$ exactly. The first few steps give

$$P(0) = 1 - p,$$ (8)

$$P(1) = p(1-q)^2e^{-2k\phi p},$$ (9)

$$P(2) = p(1-q)^2[2q + 2k\phi p(1-q)^2]e^{-4k\phi p}.$$ (10)

It is straightforward to verify that these are correct. We could also iterate Eq. (7) numerically and then estimate $P(n)$ using, for instance, forward differences at $z = 0$. Unfortunately, like many calculations involving numerical derivatives, this method suffers from severe machine-precision problems which limit us to small values of n, on the order of $n \lesssim 20$. A much better technique is to evaluate $H(z)$ around a contour in the complex plane and calculate the derivatives using the Cauchy integral formula:

$$P(n) = \frac{1}{n!}\frac{d^nH}{dz^n}\bigg|_{z=0} = \frac{1}{2\pi i}\oint\frac{H(z)}{z^{n+1}}dz.$$ (11)
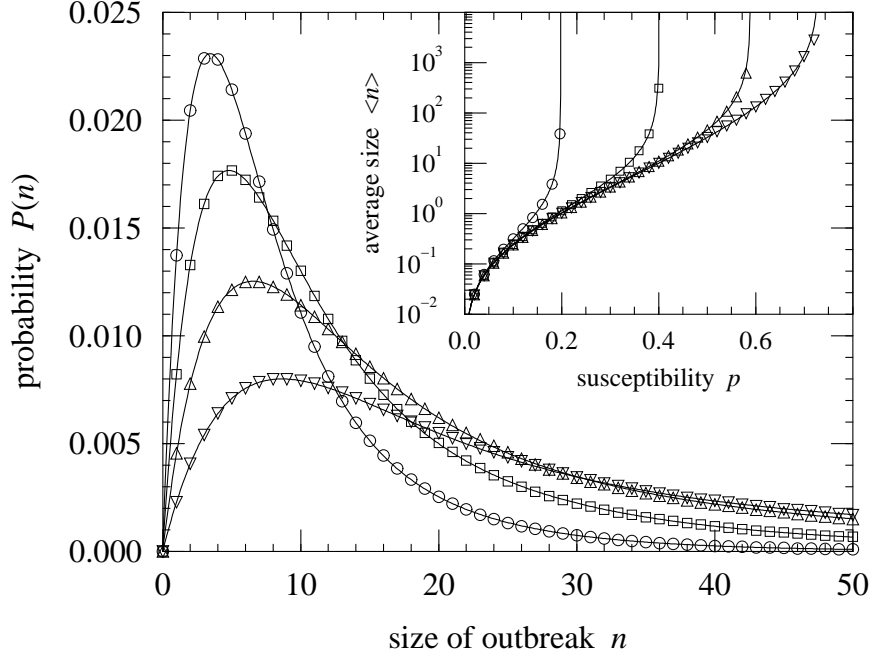
5

FIGURE 2   The distribution of outbreak sizes in simulations of the site percolation model with $L = 10^7$, $k = 5$, $\phi = 0.01$, and $p = 0.25$, 0.30, 0.35, and $p = p_c = 0.40101$ (circles, squares, and up- and down-pointed triangles respectively). The solid lines are the same distributions calculated using Eqs. (7) and (11). Inset: The average size of disease outbreaks as a function of p for (left to right) $\phi = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$. The points are numerical results for $L = 10^7$, $k = 5$ and the solid lines are the exact result, Eq. (12).

A good choice of contour in the present case is the unit circle $|z| = 1$. Using this method we have been able to calculate the first thousand derivatives of $H(z)$ without difficulty.

In Fig. 2 we show the distribution of outbreak sizes as a function of n calculated from Eq. (11) for a variety of values of p. On the same plot we also show the distribution of outbreaks measured in computer simulations of the model on systems of $L = 10^7$ sites. As the figure shows, the agreement between the two is excellent.

We can also calculate any moment of the distribution in closed form using Eq. (7). For example, the mean outbreak size is given by the first derivative of H:

$$\langle n \rangle = H'(1) = \frac{H_0'(1)}{1 - 2k\phi H_0'(1)} = \frac{p(1+q)}{1 - q - 2k\phi p(1+q)}, \tag{12}$$

6

and the variance is given by

$$\langle n^2 \rangle - \langle n \rangle^2 = H''(1) + H'(1) - [H'(1)]^2,$$
$$= \frac{p[1 + 3q - 3q^2 - q^3 - p(1-q)(1+q)^2 + 2k\phi p^2(1+q)^3]}{[1 - q - 2k\phi p(1+q)]^3}. \tag{13}$$

In the inset of Fig. 2 we show Eq. (12) for various values of ϕ along with numerical results from simulations of the model, and the two are again in good agreement.

The mean outbreak size diverges at the percolation threshold $p = p_c$. This threshold marks the onset of epidemic behavior in the model [6] and occurs at the zero of the denominator of Eq. (12). The value of p_c is thus given by

$$\phi = \frac{1 - q_c}{2kp_c(1 + q_c)} = \frac{(1 - p_c)^k}{2kp_c(2 - (1 - p_c)^k)}, \tag{14}$$

in agreement with Ref. [13]. The value of p_c calculated from this expression is shown in the left panel of Fig. 3 for three different values of k.

The denominator of Eq. (12) is analytic at $p = p_c$ and has a non-zero first derivative with respect to p, so that to leading order the divergence in $\langle n \rangle$ goes as $(p_c - p)^{-1}$ as we approach percolation. Defining a critical exponent σ in the conventional fashion $\langle n \rangle \sim (p_c - p)^{-1/\sigma}$, we then have

$$\sigma = 1. \tag{15}$$

Near p_c we expect $P(n)$ to behave as

$$P(n) \sim n^{-\tau} e^{-n/n^*} \qquad \text{as } n \to \infty. \tag{16}$$

It is straightforward to show [14] that both the typical outbreak size n^* and the exponent τ are governed by the singularity of $H(z)$ closest to the origin: $n^* = (\log z^*)^{-1}$, where z^* is the position of the singularity, and $\tau = \alpha + 1$, where $H(z) \sim (z^* - z)^\alpha$ close to z^*.

In general, the singularity of interest may be either finite or not; the order of the lowest derivative of $H(z)$ which diverges at z^* depends on the value of α. In the present case, $H(z^*)$ is finite but the first derivative diverges, and we can use this to find z^* and α.
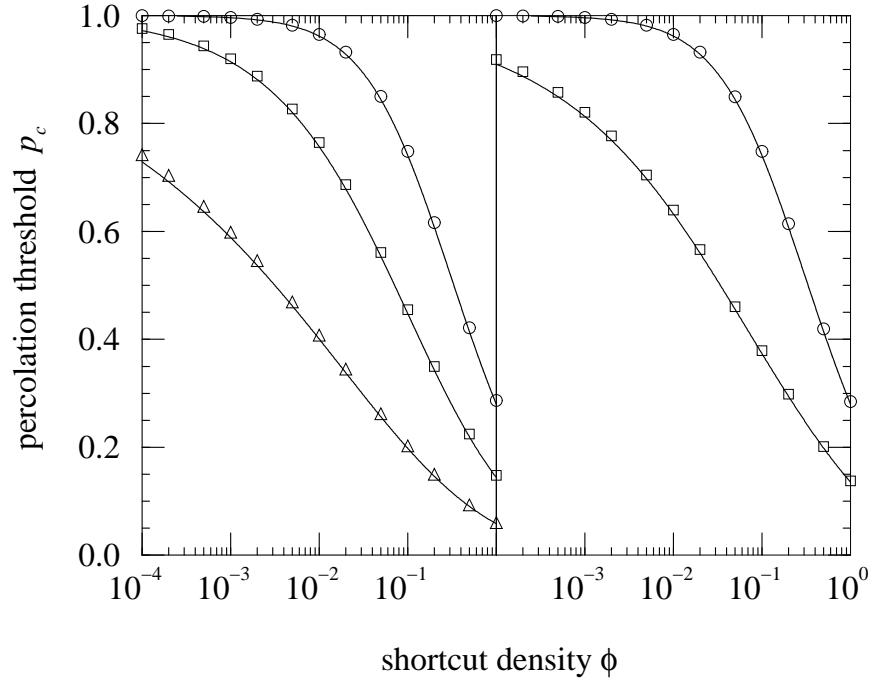
FIGURE 3 Numerical results for the percolation threshold as a function of shortcut density ϕ for systems of size $L = 10^6$ (points). Left panel: site percolation with $k = 1$ (circles), 2 (squares), and 5 (triangles). Right panel: bond percolation with $k = 1$ (circles) and 2 (squares). The solid lines are the analytic expressions for the same quantities, Eqs. (14), (25), and (26).

Although we do not have a closed-form expression for $H(z)$, it is simple to derive one for its functional inverse $H^{-1}(w)$. Putting $H(z) \to w$ and $z \to H^{-1}(w)$ in Eq. (7) and rearranging we find

$$H^{-1}(w) = H_0^{-1}(w)\, e^{2k\phi(1-w)}. \tag{17}$$

The singularity in $H(z)$ corresponds to the point w^* at which the derivative of $H^{-1}(w)$ is zero, which gives $2k\phi z^* H_0'(z^*) = 1$, making $z^* = e^{1/n^*}$ a real root of the cubic equation

$$(1 - qz)^3 - 2k\phi pz(1 - q)^2(1 + qz) = 0. \tag{18}$$

The second derivative of $H^{-1}(w)$ is non-zero at w^*, which implies that $H(z) \sim (z^* - z)^{1/2}$ and hence $\alpha = \frac{1}{2}$ and the outbreak size exponent is

8

$$\tau = \tfrac{3}{2}. \tag{19}$$

A power-law fit to the simulation data for $P(n)$ shown in Fig. 2 gives $\tau = 1.501 \pm 0.001$ in good agreement with this result.

The values $\sigma = 1$ and $\tau = \tfrac{3}{2}$ put the small-world percolation problem in the same universality class as percolation on a random graph [15], which seems reasonable since the effective dimension of the small-world model in the limit of large system size is infinite [6] just as it is for a random graph.

We close our analysis of the site percolation problem by noting that Eq. (7) is similar in structure to the equation $H(z) = ze^{H(z)}$ for the generating function of the set of rooted, labeled trees. This leads us to conjecture that it may be possible to find a closed-form expression for the coefficients of the generating function $H(z)$ using the Lagrange inversion formula [16].

Turning to bond percolation, we can apply the same formalism as above with only two modifications. First, the probability $P_0(n)$ that a site belongs to a local cluster of size n is different for bond percolation and consequently so is $H_0(z)$ (Eq. (6)). For the case $k = 1$

$$P_0(n) = np^{n-1}(1-p)^2, \tag{20}$$

where p is now the bond occupation probability. This expression is the same as Eq. (3) for the site percolation case except that $P_0(0)$ is now zero and $P_0(n \geq 1)$ contains one less factor of p. $H_0(z)$ for $k = 1$ is

$$H_0(z) = z\frac{(1-p)^2}{(1-pz)^2}. \tag{21}$$

For $k > 1$, calculating $P_0(n)$ is considerably more complex, and in fact it is not clear whether a closed-form solution exists. However, it is possible to write down the form of $H_0(z)$ directly using the method given in Ref. [13]. For $k = 2$, for instance,

$$H_0(z) = \frac{z(1-p)^4(1-2pz+p^3(1-z)z+p^2z^2)}{1-4pz+p^5(2-3z)z^2-p^6(1-z)z^2+p^4z^2(1+3z)+p^2z(4+3z)-p^3z(1+5z+z^2)}. \tag{22}$$

9

The second modification to the method is that in order to connect two local clusters a shortcut now must not only be present (which happens with probability ϕ) but must also be occupied (which happens with probability p). This means that every former occurrence of ϕ is replaced with ϕp. The rest of the analysis follows through as before and we find that $H(z)$ satisfies the recurrence relation

$$H(z) = H_0(ze^{2k\phi p(H(z)-1)}),\tag{23}$$

with H_0 as above. Thus, for example, the mean outbreak size is now

$$\langle n \rangle = H'(1) = \frac{H_0'(1)}{1 - 2k\phi pH_0'(1)},\tag{24}$$

and the percolation transition occurs at $2k\phi pH_0'(1) = 1$, which gives

$$\phi = \frac{1 - p_c}{2p_c(1 + p_c)}\tag{25}$$

for $k = 1$ and

$$\phi = \frac{(1 - p_c)^3(1 - p_c + p_c^2)}{4p_c(1 + 3p_c^2 - 3p_c^3 - 2p_c^4 + 5p_c^5 - 2p_c^6)}\tag{26}$$

for $k = 2$. As in the site percolation case, the critical exponents are $\sigma = 1$ and $\tau = \frac{3}{2}$. In the right panel of Fig. 3 we show curves of p_c as a function of ϕ for the bond percolation model for $k = 1$ and $k = 2$, along with numerical results for the same quantities. The agreement between the exact solution and the simulation results is good.

We can also apply our method to the case of simultaneous site and bond percolation, by replacing $P_0(n)$ with the appropriate distribution of local cluster sizes and making the replacement $\phi \to \phi p_{\mathrm{bond}}$ as above. The developments are simple for the case $k = 1$ but the combinatorics become tedious for larger k and so we leave these calculations to the interested (and ambitious) reader.

To conclude, we have studied the site and bond percolation problems in the Watts–Strogatz small-world model as a simple model of the spread of disease. Using a generating function method we have calculated exactly the position of the percolation transition at

which epidemics first appear, the values of the two critical exponents describing this transition, and the sizes of disease outbreaks below the transition. We have confirmed our results with extensive computer simulations of disease spread in small-world networks.

Finally, we would like to point out that the method described here can in principle be extended to small-world networks built on underlying lattices of higher dimensions [2,6]. Only the generating function for the local clusters $H_0(z)$ needs to be recalculated, although this is no trivial task; such a calculation for a square lattice with $k = 1$ would be equivalent to a solution of the normal percolation problem on such a lattice, something which has not yet been achieved. Even without a knowledge of $H_0(z)$, however, it is possible to deduce some results. For example, we believe that the critical exponents will take the values $\sigma = 1$ and $\tau = \frac{3}{2}$, just as in the one-dimensional case, for the exact same reasons. It would be possible to test this conjecture numerically.

REFERENCES

[1] S. Milgram, "The small world problem," Psychology Today **2**, 60–67 (1967).

[2] D. J. Watts, *Small Worlds.* Princeton University Press, Princeton (1999).

[3] B. Bollobás, *Random Graphs,* Academic Press, New York (1985).

[4] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature* **393**, 440–442 (1998).

[5] M. Barthélémy and L. A. N. Amaral, "Small-world networks: Evidence for a crossover picture," Phys. Rev. Lett. **82**, 3180–3183 (1999).

[6] M. E. J. Newman and D. J. Watts, "Scaling and percolation in the small-world network model," Phys. Rev. E **60**, 7332–7342 (1999).

[7] C. F. Moukarzel, "Spreading and shortest paths in systems with sparse long-range connections," Phys. Rev. E **60**, 6263–6266 (1999).

[8] R. V. Kulkarni, E. Almaas, and D. Stroud, "Exact results and scaling properties of small-world networks," Phys. Rev. E, in press. Also `cond-mat/9908216`.

[9] M. E. J. Newman, C. Moore, and D. J. Watts, "Mean-field solution of the small-world network model," Santa Fe Institute working paper 99–09–066. Also `cond-mat/9909165`.

[10] R. Monasson, "Diffusion, localization and dispersion relations on small-world lattices," Eur. Phys. J. B **12**, 555–567 (1999).

[11] A. Barrat and M. Weigt, "On the properties of small-world network models," Eur. Phys. J. B **13**, 547–560 (2000).

[12] M. E. J. Newman, "Small worlds: The structure of social networks," Complexity, in press (2000). Also `cond-mat/0001118`.

[13] C. Moore and M. E. J. Newman, "Epidemics and percolation in small-world networks,"

Santa Fe Institute working paper 00–01–002. Also `cond-mat/9911492`.

[14] G. H. Hardy and J. E. Littlewood, "Tauberian theorems concerning power series and Dirichlet's series whose coefficients are positive," Proc. London Math. Soc. **13**, 174–191 (1914).

[15] D. Stauffer and A. Aharony, *Introduction to Percolation Theory,* 2nd edition. Taylor and Francis, London (1992).

[16] H. S. Wilf, *Generatingfunctionology,* 2nd Edition. Academic Press, London (1994).